

### 3. MODEL CONCEPTUAL DESIGN

This chapter presents a conceptual design of the GIS-based automobile exhaust emissions model. The background research from the previous chapter is summarized into a series of ‘research foundation points’ that define modeling parameters. User requirements are also identified, guiding model form and presentation. By the end of the chapter, a modeling approach is recommended.

While the overall purpose of the model is defined in Chapter 1, more specific model objectives that guide the development of such a model include:

- *The model must produce automobile exhaust emission estimates that are capable of being statistically verified.*

It is vital for the model to be able to determine errors in estimates that result from input data error and algorithm error. One of the biggest criticisms of the currently mandated modeling approach is that there is no information available for users to estimate errors resulting from the algorithms. A design open to outside review and analysis prevents avoidable extrapolation because the confidence intervals are known.

- *All estimates and input parameters (emissions, vehicle activity, etc.) must be capable of being validated.*

All model components must be capable of being validated either through previously published research or through designed experiments. Given the complicated process of predicting emissions, it is important that all intermediate modeling steps be designed to be tested. This objective will influence the data model because many elements regarding vehicle technology and vehicle activity will have to have identifiable characteristics that can be observed in the field.

- *The model must be designed to easily incorporate new findings.*

Because research into emissions modeling is occurring in a number of institutions, significant findings are expected in the near future. Keeping the research model up to date to research from other institutions is crucial if it is going to be used to influence research direction and software development decisions.

- *The model must use available data.*

Although the model is not designed to be implemented on a widescale basis for official reporting, it must still be constrained by real-world conditions of data availability and cost. Without considering these factors, one can spend a significant

amount of time and resources developing models from variables (dynamic engine parameters, etc.) that cannot be collected by a regional modeling agency.

- *The model must use as large a spatial scale as data will allow.*

It is important to use available data, but it is also important to use the largest scale possible. One of the uses of the model will be to identify the level of spatial aggregation required for useful emission estimation. In order to do this, it is important to start with the most detail and aggregate up, thereby identifying locations with high emission production.

- *The model must be portable.*

The model should be transferable to other urban areas without substantial model alteration. This means that all the input parameters should be available to major metropolitan areas, and that model assumptions should not be limited to the study area.

### **3.1. Model Design Parameters**

The following model design parameters are based on material discussed in the background chapter, user requirements, and good modeling practice. These parameters will establish minimally acceptable guidelines for model development. The ability of the model design to abide by the parameters will depend on the data and technology available to a clearly-defined user group. Some parameters may have to be scaled back due to limitations in data availability.

This summary of the background knowledge is presented in this section to clearly identify model development parameters. The initial goal of model development is to include all listed parameters. At some point limited data availability or excessive data development expense will likely remove or scale back some parameters from consideration. The research backed parameters are:

- *Develop estimates of the production of automobile exhaust pollutants CO, HC, and NOx in space and time (from section 2.1.1)*

Research has shown that the major exhaust pollutants of concern are CO, HC, and NOx. Considerations should be given to including particulate matter greater than 2.5 microns in diameter (PM2.5) due to its recent identification as a health risk. However, there is very little data on the cause and effect relationships of PM2.5 production by automobiles. The emission estimates represent only the production of pollutants, not the resulting air quality. The spatial and temporal scale should be developed according to anticipated user needs. Existing photochemical models (used to predict ambient air quality) currently use hourly, 4-5 sq. km aggregations. Future

photochemical model improvements are expected to use 1 sq. km estimates of mobile sources.

- *Anthropogenic NO<sub>x</sub> estimate accuracy important in predicting ground-level ozone (from section 2.1.1)*

Major cities in warmer climates have air quality problems resulting from ground level ozone concentrations. NO<sub>x</sub> and HC are precursors to ozone formation. HC, however, can be produced in significant amounts by biogenic sources. Therefore, a more accurate, verifiable, estimate of NO<sub>x</sub> may prove more useful in predicting the impact of motor vehicles.

- *Comprehensive representation of vehicle technologies (from section 2.1.2)*

Differences in vehicle technologies / characteristics have been shown to significantly affect vehicle emission rates. As seen in the physical model approach by Barth, et al., it is actually the dynamic status of a number of vehicle parameters that causes emission rate variability (see section 2.2.2). At the same time, a number of vehicle characteristics have been tied to emission rate variability because they are surrogate variables for causal parameters (see section 2.2.3). From a research model perspective, it is important to be able to include both sets of conditions. However, modeling individual vehicle engine dynamics for an urban area is not practical due to extensive data requirements. Instead, only those specific static inventory variables involved with the dynamic conditions, and those variables identified as surrogate variables are included. The list of desired vehicle characteristics are: model year, engine size, weight (or mass), emission control type(s), fuel delivery type, transmission type, cross-sectional area, and number of cylinders.

- *Separate and quantify high-emitting vehicle emissions (from section 2.1.2)*

A small percentage of the fleet disproportionately contributes to total mobile source emissions. By separating this small high-emitting portion of the operating fleet, it will be easier to predict the impacts of control strategies that may target high emitters. Further, model attention should be focused on factors that result in higher emissions, wisely using resources in the most important areas.

- *Separate start, hot-stabilized, and enrichment emission quantities and locations (from section 2.1.3)*

By separating estimates into specific emission modes, mode-specific impact strategies can be more efficiently evaluated. Further, emission rates for each mode are predicted using different variables. Engine starts are primarily influenced by vehicle characteristics and engine temperature. Hot-stabilized and enrichment emissions are primarily influenced by vehicle characteristics and operating condition (speed, acceleration, etc.).

- *Include Speed Correction Factor (SCF) emission rates (from section 2.2.1)*

The inclusion of SCF emission rates provides an alternative modeling approach. One of the objectives of this research is to be comprehensive and flexible. Inclusion of the SCF estimate provides a flexible framework and a way to compare between emission rate modeling approaches. The model may indicate that the highly aggregate SCF approach is suitable for regional inventory modeling at a certain spatial level. The SCF approach to modeling start emissions will not be included because the approach lacks the ability to show spatial variability between start and running emissions.

- *Include emission rates from the statistical approach (from section 2.2.3)*

Emission rates from the statistical approach need to be included because the research indicates that modal parameters better characterize accurate emission rate estimation. Because the modal emission rates models are available, they can be immediately integrated into the model framework. The approach also produces separate start and running exhaust emission estimates, addressing one of the previously defined model design parameters.

- *Include activity measures from travel demand forecasting models (from section 2.3.1)*

Travel demand forecasting models are the primary predictive tool for regional level vehicle activity. They are also used by almost every transportation planning agency (MPO) in the country. Further, their use in developing emission estimates is currently mandated. Despite their well-documented problems, they have characteristics that make them very attractive for a spatially-resolved model. First of all, they have a defined structure and connectivity that translates into a spatial form (zones, links, and nodes). Second, they develop estimates using socioeconomic information, allowing the model to be indirectly affected by social and economic changes.

- *Prepare for inputs from future simulation models (from section 2.3.2)*

Simulation models provide vehicle activity measures at a larger spatial scale and resolution than macroscopic travel demand models. The value of producing estimates at the microscopic or mesoscopic scale becomes evident when studying the types of vehicle activity that produce high emissions. Just as high emitters disproportionately contribute to total emissions, so do high power demand situations. These driving situations can be characterized by comprehensive representation of traffic flow dynamics.

- *Use geographic information systems (from section 2.4)*

Using GIS is important because it is designed to handle the spatial data management and modeling functions key to the research goals. Without GIS, complex spatial analysis and manipulation algorithms would have to be re-created. Its widespread use and popularity among planning agencies is significant enough to warrant its use.

## **3.2. User Requirements**

In designing any analysis model, it is crucial to clearly understand the analysis needs of the proposed users. There are several user groups that could be expected to interact with the research model.

- *Emission Science Experts*

These experts are those individuals who help define the emission science domain. They provide the knowledge regarding the cause and effect relationships in automobile emissions modeling. Although their interaction with model design is conceptual, it is tremendously beneficial if they can interact with specific model components to ensure that the science is being accurately represented. Therefore, one data model requirement is that the system be composed of well-documented and appropriately termed modules that can be easily reviewed by the specific component's experts. The model vocabulary should be defined by the experts' terminology (i.e., transportation components use standard traffic engineering terminology).

- *Model Developers*

Model developers can also have significant knowledge of the cause and effect relationships among key variables. If a model requires significant software development, it would be prudent to organize it using standard programming techniques, terminology, and comments. This may require that comments in the code explain the underlying scientific concepts to the point that clear understanding of the importance of the various pieces is evident to developers. If a developer could improve program efficiency by slightly altering a process, it would be beneficial that the explanatory cost of the change be evident. Therefore, well-documented code is a specific data model requirement.

- *Emission Researchers*

Emission researchers are individuals who use the model to get a better understanding of the impacts of new findings, or develop criteria for future research efforts. This adds a dimension to the model by requiring that measures of confidence be included with the estimates. The estimates produced by the model must be capable of being accompanied with certain measures of accuracy and descriptions that clearly identify what is known and unknown in the process. The model inputs and outputs should be in a format that allows easy import/export to various software packages that may be used for more detailed analysis. Outputs must include detailed summaries of assumptions and discussions of accuracy to prevent false conclusions from being drawn.

- *Government Experts*

Government experts would be individuals who would look at all levels of model development to ensure quality and accuracy in order to approve or disapprove results that could hold legal bearing. If model results are to be used for conformity or inventory reporting, the model elements must be validated and peer-reviewed. This is important in a developing model because government experts and researchers must be included in the design process to prevent efforts from moving in directions that contradict policies and mandates that govern air quality modeling. This user requirement strengthens the need for modular, clearly communicated model code.

- *Transportation and Environmental Planners*

Transportation and environmental planners are the eventual ‘users’ of the system. They will be the ones who develop the emission estimates for their particular project. Although the level of development discussed in this report is for a ‘research-grade’ model not to be used for legal reporting, the intention is that the model or some of its components eventually be targeted for widespread public use. By including the eventual user needs in the early design, complications in future development can be avoided. By including transportation and environmental professionals, who may or may not have model development experience, the system design becomes intuitive and flexible. Planners should not be burdened with extensive command and syntax requirements. Results should be designed towards the reporting needs of the planners.

- *Non-technical Decision-Makers*

Decision-makers (policy-makers, managers, planning boards, etc.) need model results to make informed decisions. Decisions range from guiding the direction of research to broad-based policy analysis and to local transportation alternative analyses. Many of these users are removed from the modeling process, but must be familiar with the process of modeling so they can have confidence in the model results and be aware of assumptions made by modelers. By maintaining the model framework within an off-the-shelf GIS, questions about model inputs and outputs can be asked and answered by non-technical users. Further, thematic maps and user-defined spatial queries, graphical results can be produced along with standard spreadsheet and textual reports.

In summary, the user-defined needs include:

- *Appropriate documentation*
- *Appropriate terminology*
- *Modular system design*
- *Open input and output data formats*
- *Intuitive modeling process*
- *Easy to understand and use*
- *Model should reside in a GIS*

### **3.3. The Spatial Data Model**

A good data model has five dimensions [Reingruber, et al., 1994]: conceptual correctness, conceptual completeness, syntactic correctness, syntactic completeness, and enterprise awareness. *Conceptual correctness* refers to the degree to which the model represents the real world, or the accuracy of the model estimates. For this model, it refers to the accurate representation of the cause and effect relationship between motor vehicle behavior and emissions. *Conceptual completeness* refers to the wholeness of the represented science. In this model, it refers to the ability to represent the cause and effect relationships in a comprehensive manner for the entire urban area. *Syntactic correctness and completeness* refer to the quality of the use of language and proper communication. This would involve the use of organized and structured programming techniques as well as the use of accepted transportation, air quality, and GIS terminology. *Enterprise awareness* refers to the idea that the model does not work in a vacuum, but that it represents only a portion of a much larger system. An automobile exhaust model is a portion of the much larger scope of environmental and transportation modeling. Keeping the model open to connectivity with other systems ensures an adaptable and open system.

As the design of model components develops, it will be analyzed in respect to these five concepts, paying particular attention to completeness and correctness. These measures will guarantee that the design will have an organized framework.

Spatial data entities are the spatial forms used to characterize an object. For example, a road ‘object’ can be characterized by the digital representation of a line, the spatial data entity. Generally, the best entity type to use for spatial modeling or environmental data representation is a raster cell. A raster cell structure handles continuous variables better than a vector (points, lines, and polygons) structure. This occurs because regular grid cells that fall between observed values can have statistically interpolated values. A vector representation forces observed values to be discrete within its structure, possibly misrepresenting boundaries. Another possible spatial data entity is a triangulated irregular network (TIN). TINs interpolate points found on a line between two observation points based on the values of the points. It is most often used in representing topography, but the concept could be translated to other areas. The selection of an appropriate spatial structure is controlled by the model objectives, model parameters, and user needs.

A vector approach has the following advantages:

- *Intermediate estimates must be validated*

Because the model being developed is research-oriented, all of the algorithms and data must be represented in a format that can be field validated. A clearly defined

beginning and ending point for a segment of road (intersection to intersection), or a field-evident boundary (zone bounded by roads) makes validation simpler.

- *Users require facility-level estimates*

Transportation modelers (eventual users of the model) work with vector-based facility entities (see section 3.4.1). By providing and receiving data in a similar format, the integration and transfer of data is more efficient.

- *Pollutant production is discrete*

Vehicles are discrete objects. Because the model predicts pollutant production, not resulting air quality, the emission estimate should also be discrete. Given that the model will not actually model individual vehicles, there is an argument that an aggregate characterization of factors is more efficiently handled in a raster approach. However, the appropriate level of aggregation is undefined; in fact, one of the model objectives is to provide a way to determine the appropriate level of aggregation. Once it is determined, a raster approach for software development may be warranted.

A raster approach to the model also has benefits associated with its use in the research model:

- *Inventory estimates are gridded*

The final outputs of mobile emissions models are inputs into photochemical models. The photochemical models are raster due to the nature of the phenomenon being modeled (ambient air quality). Gridded, hourly estimates are currently required.

- *Data will have to be aggregated eventually*

It is unlikely that widespread data availability allows modeling on a vehicle per vehicle basis, nor is it likely that modeling at that level is practical or useful for regional scale modeling. Some level of aggregation will have to be used. In that regard, it may become more appropriate to predict continuous distributions rather than discrete polygon values.

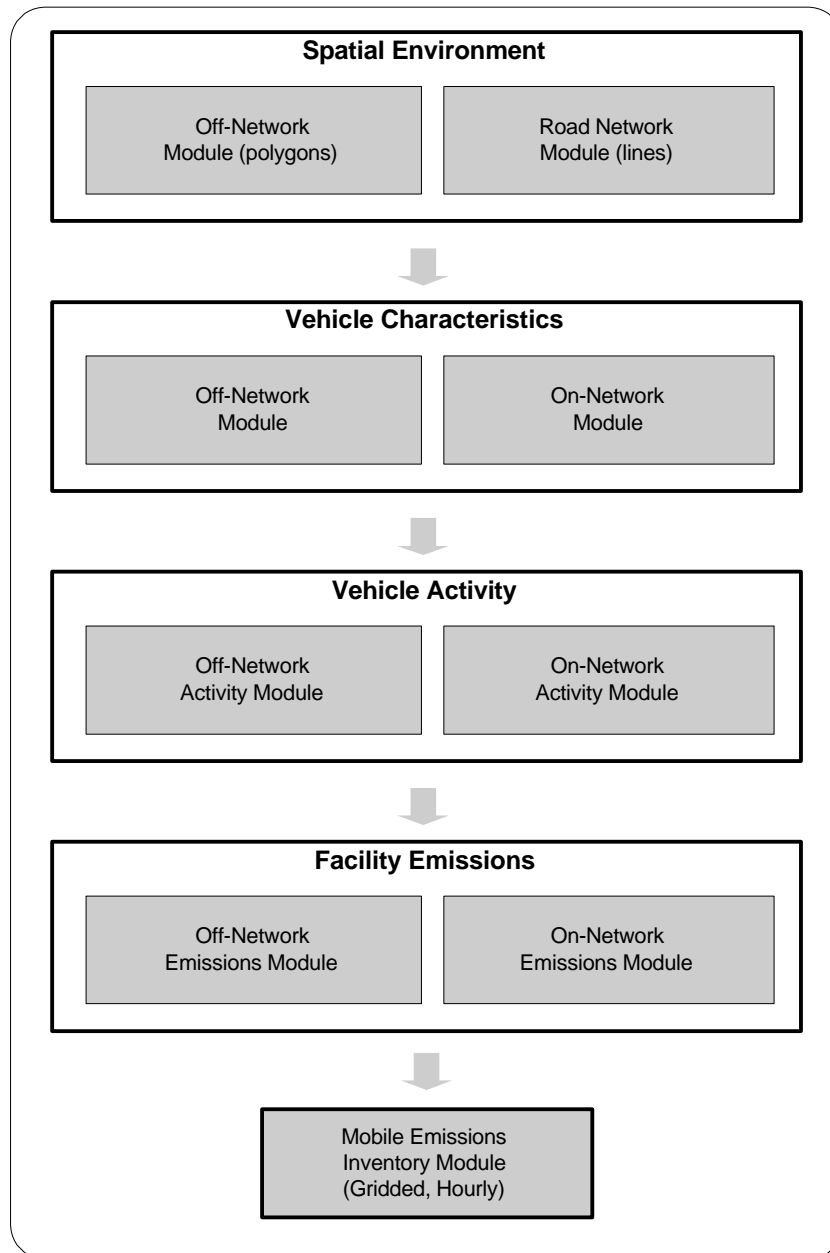
Overall, if the user is concerned with the location of high auto emissions, a raster approach would be better, although technically questionable for linear features such as roads. If the user is concerned with the emissions on a specific entity (i.e., road segment, TAZ), a vector approach would be better. Given that the specified users are concerned with both issues, both raster and vector entities should be used. The issues of validation and emission science suggest that initially, vector data models are warranted. At some point, the vector structure needs to be converted to raster for further photochemical modeling, and regional data visualization.

### **3.4. Model Approach**

The conceptual design of the proposed research model meets all of the stated goals and objectives that have been stated earlier while avoiding the constraints of extensive data development and cost. The model will be deterministic and spatially and temporally identify: the types of vehicles that are being operated, the types of activities the vehicles are involved in, the resulting emission rates, and the resulting emissions. The level of aggregation and spatial scale is flexible, depending upon the user's needs, data availability, and accuracy requirements of post-processing. Regardless of the spatial scale, the conceptual design remains the same. Figure 3.1 shows a schematic drawing of the design. The top row represents the spatial environment. The second represents vehicle characteristic assessment. The third represents the vehicle activity. The fourth and fifth rows represent facility-level and gridded emission estimation. Detailed information about each component is provided in chapter 4.

Central to the model design is the identification of the source of vehicle activity data. While travel demand forecasting models have significant limitations providing inputs to emissions modeling, they represent the only widely used prognostic planning tool available. Until regional micro-simulation models become widely accepted, validated, and implemented, emission models must rely on the forecasting capabilities of the tools in use. With all of its disadvantages, there are components of the travel demand forecasting model that defend its use for emissions modeling. First of all, trip generation results can be easily translated to engine starts, an important emission activity. Second, poor estimates of average speed can be supplemented with observed speed and acceleration data given certain traffic flow parameters. Third, the travel models have spatial characteristics that can form a foundation for spatial modeling. The research model is tied to traditional travel demand models. At some point in the future, regional simulation models may become the primary source of travel behavior prediction. This should be reflected in the model design by avoiding strategies that lock into specific travel model types.

**Figure 3.1 - Conceptual Model Design**



The following sections describe the five major tiers of the model design. In each section, descriptions of the roles, data needs, and processes are provided. Three of the five parts of a good data model are discussed for each major component: conceptual correctness, conceptual completeness, and enterprise awareness (the other two parts are related to syntax and deemed less significant). Each section is supplemented with specific model descriptions in chapter 4.

### 3.4.1. Spatial Environment

The objective of the spatial environment tier is to unify input data under a common zonal and lineal structure. The size and scope of the zones and lines depend on the users and their specific needs. Historically, exhaust emissions are divided into

start and non-start (running exhaust) emissions. Most prognostic travel models provide a zonal (TAZ) estimate of the number of trip origins, and a lineal (link) estimate of road volume and average speed. By defining an engine start as being synonymous with a trip origin, TAZs become the base spatial entity used for engine start emissions. Running exhaust emissions occur on the road network, suggesting that a ‘link’ should be the base spatial entity. Improvements in the spatial resolution of the zonal estimates can be made outside the travel model by disaggregating trips to smaller zones. The lineal estimate can similarly be improved by conflating (see section 3.4.1.3) the links to comprehensive and accurate road datasets. These issues are discussed further in the next two sections.

#### **3.4.1.1. Zonal Data**

The zonal module defines the polygon structure used to represent data and emission estimates for engine starts. It is the role of the zonal module to combine the polygons of various input data (i.e. socioeconomic, land use, TAZ) into a single polygon dataset. As mentioned before, TAZs represent the base spatial entity for engine starts. However, disaggregating trip origin estimates from large TAZs to smaller zones can be accomplished if good socioeconomic and land use data are available. For example, home to work trip origins can be assumed to start from the residential areas within the TAZ. Likewise, return trip origins can be assumed to start from land uses representing workplaces. While the process of disaggregation is discussed later, it is the role of the zonal module to establish the data linkages that make it possible.

Due to the fact that polygon data usually come from a variety of original sources and therefore a variety of spatial representation differences, significant errors may occur when trying to bring the datasets into a unified structure. It is unlikely that boundaries that represent identical features from different datasets will match perfectly. The result of this problem is the creation of a series of ‘sliver’ polygons whose attributes may be misaligned. However, there is no loss of information, only a zone structure that is as spatially accurate as the original data. The model does not make any assessment about the spatial accuracy, but uses whatever data are available. This allows users to define their spatial accuracy needs through the accuracy of the input data. Thus, if one wants estimates of engine start pollutant production within 100 meters, one must provide input data with equal or better spatial accuracy than 100 meters. Each of the new polygons maintains key fields tying them to their original datasets, allowing all engine start emission estimates to be aggregated to any of the input polygon structures.

#### **3.4.1.2. Lineal Data**

The road module defines the lineal data used for predicting running exhaust emissions. While the travel demand forecasting models continue to be criticized for inaccurate roadway volume and speed estimates, they represent the only available prognostic regional vehicle activity tool. Most of the models predict travel (volume and average speed) only for major roads, aggregating minor roads to TAZ ‘centroid connectors’. Further, the lineal representations of the road networks are usually spatially abstract structures. ‘Links’ represent actual road segments, but given modeling tasks, detailed shape points are unnecessary. In Atlanta, the absolute spatial errors resulting from the abstract representation exceeded 2 km in some instances [Bachman, 1996]. Improving this error is important to generating emissions for grid cells of 4 sq. km or smaller.

#### **3.4.1.3. Conflation**

Conflation is the blending of two line databases. Conflating the abstract travel demand forecasting network and a spatially accurate comprehensive road database is needed to improve the spatial accuracy of the travel model results. Because travel models’ abstract ‘links’ represent actual road segments, it is possible to assess the connections to other road dataset lines based on link configuration and attributes. The process requires a link by link assessment and conflation by the user, resulting in a time-consuming and tedious task. Many planning organizations that develop travel demand forecasting models have already conflated the networks for purposes outside emission modeling. Conflation is required for the research model and not considered to be a task beyond the users needs. The model design can function without it, but at a significant loss of spatial error, thereby eliminating one advantage of the approach.

The purpose of the model’s road module is to separate the conflated road dataset into modeled and unmodeled roads. Modeled roads, usually the roads with the most volume, become the major lineal structure used to represent running exhaust emissions. Unmodeled roads are aggregated into zones (bounded by modeled roads) used to represent minor running exhaust emissions. An argument that supports the zonal representation is the assumption that half the vehicles traveling on minor roads have a higher chance of operating under start conditions because they are closer to their origin (start conditions can last 2-3 minutes).

- *Spatial Environment Conceptual Correctness*

The spatial environment modules have the task of defining the locational parameters for the rest of the model. The structure of the spatial environment needs to reflect the spatial characteristics of automobile exhaust emission production.

Automobile exhaust emissions are produced by operating vehicles traversing a road network. The road network becomes the crucial component of the spatial environment. For major roads, there is little loss in the conceptual correctness of the spatial representation. Minor roads, however, suffer from insufficient prognostic data forcing zonal aggregation. The zonal aggregation uses discrete polygons in representing urban information. While the conceptual correctness suffers, important linkages become straightforward and the needs for minor road modal activity are lessened.

- *Spatial Environment Conceptual Completeness*

The conceptual completeness of the spatial environment refers to the comprehensiveness of the spatial representation. The zonal aggregation of all roads not modeled by the travel demand forecasting model ensures comprehensive spatial representation by being a ‘catch-all’ for minor roads. The age of the input data will impact the completeness of the data. Recent land use changes or new road construction will be left out unless the input data are continuously maintained.

- *Spatial Environment Enterprise Awareness*

The spatial environment structure is based on zonal and lineal representations of spatial structures used in a variety of agencies. By maintaining connections to the original input dataset identifiers, solid linkages to these agencies are provided. Further, by using a GIS and organizing data based on location, an indirect linkage to many enterprises is possible.

### **3.4.2. Fleet Characteristics**

Although different emission modeling approaches are being developed, all research efforts indicate that an improved capability to identify the emission significant components of the operating fleet is important to emission rate accuracy [Siwek, 1997]. Currently, emission models use model year distributions to describe the fleet. However, many other vehicle characteristics hold significant explanatory capability for predicting emission rates [Guensler 1994, Barth 1996]. Further, spatially variant emission estimates are needed, requiring spatially resolved sub-fleet characterization [Bachman, 1996]. Therefore, there is a need for identifying procedures that can accurately predict spatially resolved vehicle characteristics for urban areas. The fleet characteristics modules described in this section develop emission-specific and location-specific estimates of the distribution of automobiles.

Regional vehicle registration data provide information that allow emission-important vehicle characteristics to be determined for individual vehicles. The data also provide clues to identifying the vehicle’s location, the owner’s registered address, and the owner’s ZIP code. The fleet characteristics tier will develop estimates of

technology distributions for each of the zonal and lineal representations. There are four general tasks: (1) attaching location parameters to the individual vehicle registration data; (2) determining important characteristics of the vehicles; (3) determining technology groups for the vehicles, and (4); aggregating to spatial entity-specific technology distributions.

The first vehicle characteristics module has two major tasks: determine individual vehicle location parameters and emission-specific characteristics. Each task is time-consuming due to the size of the registration datasets found in a metropolitan area. In Atlanta in 1995, 2.2 million vehicles were registered in the nonattainment area. The initial intense processing tasks need to be completed only one time per year, following new registration database development. Therefore, the first vehicle characteristics module becomes a ‘pre-processing’ step, residing outside the formal model.

#### **3.4.2.1. Vehicle Geocoding**

Address geocoding is a process whereby standard address fields of road name, road type, and ZIP code are used to identify corresponding lines in a road database. The address number is used to identify the position of the address on a matched line based on the left and right address ranges. Address-matching usually results in success ranges of 60-80% dependent on the quality and comprehensiveness of the road dataset, and the number of errors associated with miscoding, duplicate or multiple road names, apartment numbers, and rural route identification. Growing urban areas have difficulty keeping road datasets current with new housing developments, adding significant bias to the geocoding errors.

The geocoding process results in two types of records: matched and unmatched. The matched vehicles are associated with a point entity. The unmatched vehicles maintain a default location identifier of ZIP code, a polygon entity. While ZIP codes can be rather large, they provide a degree of spatial information that can help determine regional fleet distribution variability.

#### **3.4.2.2. VIN Decoding**

Raw registration data can usually provide a few important vehicle characteristics (VIN, make, model, model year, and number of cylinders), but more information can be developed from the vehicle identification number (VIN). All vehicles after 1980 are given a 17 digit VIN that consists of a code containing information about the types of emission control systems, the fuel delivery systems, the engine size, etc. Prior to 1980, VINs existed but lacked universal standards. Decoding the VIN for each vehicle requires the use of software (VIN decoder) developed by

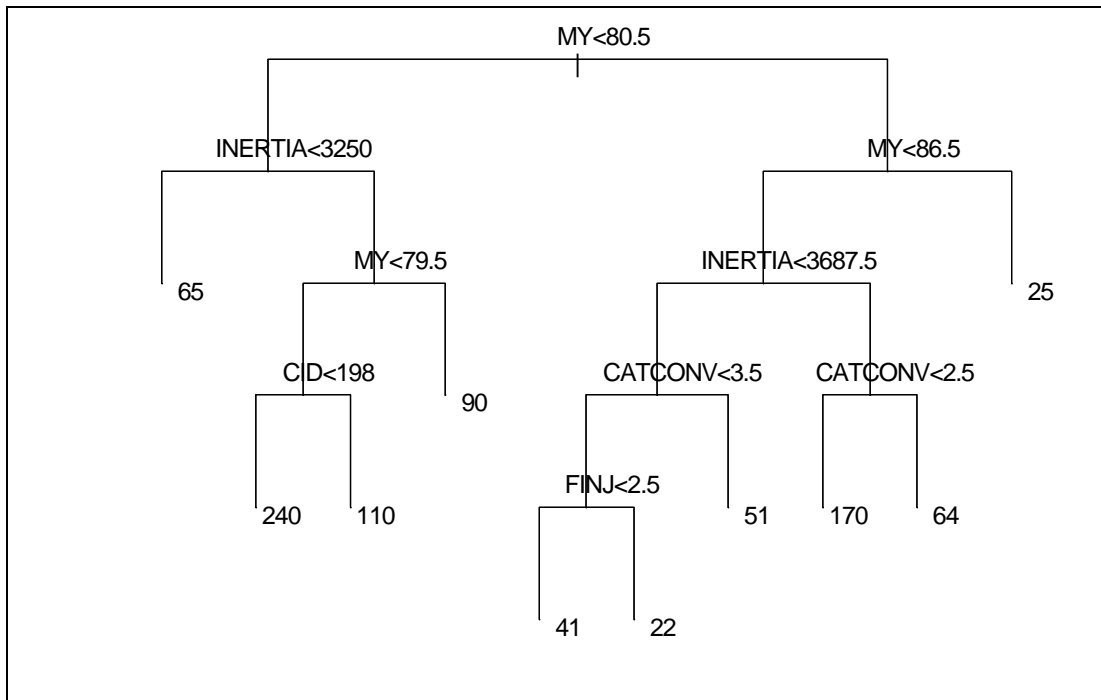
Radian International Corporation. Missing vehicle characteristics and the lack of updates prevent sole reliance on Radian's VIN decoder [Bachman, 1998]. Missing characteristics, pre-1972 autos, and post-1994 autos need to be developed from lookup files using the make, model, and model year. Research efforts at Georgia Tech have resulted in a datafile that can be used to determine the test weight of vehicles. While significant errors are expected, enough information should be available to develop a clear view of the operating fleet distributions.

The vehicle characteristics module results in two groups of vehicles and their emission specific characteristics; point-based (successfully matched) and zone-based (unmatched). These files should represent a comprehensive description of the region's fleet characteristics. These files are further processed to develop the emission-rate specific fleet distributions.

The zonal technology group module takes the spatially-resolved vehicle characteristics' files and determines zone-based engine start and running exhaust technology group distributions. The technology group (TG) definitions are defined by the emission rate modeling approaches included in the system. Currently, they are the aggregate modal approach (see section 2.2.3) and the speed correction factor approach (see section 2.2.1) because they are the only currently available models.

#### **3.4.2.3. *High and Normal Emitters***

The aggregate modal approach developed emission-specific technology groups using a regression-tree analysis of emission test vehicles [Wolf, 1998]. In the analysis, all vehicles are divided into technology classes, indicating high or normal emitter fraction likelihoods. High and normal technology groups are then defined for each pollutant and emission mode (engine start and running exhaust). A sample engine start, normal emitter, CO, 'tree' is provided in figure 3.2. By starting at the top of the tree, conditions are identified based on the vehicles characteristics. True statements move to the left side of the tree, false statements move to the right. Each ending node is a set of conditions that are assigned a grams per start emission rate.



**Figure 3.2 - Sample Regression Tree for Normal CO Engine Starts (grams/start)**

A high emitting vehicle is one that has malfunctioning or tampered with emission control systems causing higher than normal emissions. It is expected that a small percentage of high emitting vehicles account for a large percentage of total emissions. High emitter determination is an important model design parameter and therefore it is appropriate to characterize these vehicles differently. The fraction of high emitters in the fleet, and the rate of malfunction among different vehicle types are unknown, but currently being researched. The regression-tree results by Wolf et al. divide the fleet into four groups that have different likelihoods of being high emitters. Lacking better information, a random sample for each group will be separated and labeled as high emitters, with sample sizes based on the group's likelihood. All other vehicles will be modeled as normal emitters.

#### **3.4.2.4. Technology Groups**

Once vehicles are identified as high or normal emitters, they are characterized into technology groups. Technology groups are combinations of vehicle characteristics and operating conditions that have been identified in the regression tree analysis as having significant emission rate differences. There will be separate technology groups for high and normal emitters, each pollutant, and each operating mode. Each vehicle will fall into six technology groups (engine start CO, HC, and NOx, and running

exhaust CO, HC, and NO<sub>x</sub>). Specific technology group descriptions are provided in chapter 4.

Engine start technology groups only include vehicle characteristics. For each emission-significant combination of vehicle characteristics, an associated gram per start emission rate is identified. Running exhaust technology groups include vehicle characteristics and(or) modal operating parameters (idle, cruise, acceleration, etc.). Unlike engine start groups, running exhaust technology groups can have different emission rates based on modal operating conditions. By the end of the fleet characteristics' modules, distributions of technology groups will exist for every zone in the model.

- *Fleet Characteristics Conceptual Correctness*

The representation of vehicles into emission-specific high and normal emitter technology groups is based on observed relationships discovered through test datasets. The ability of the technology groups to correctly represent the emission-specific characteristics of the operating fleet directly relates to the representativeness of the emission test dataset. Clearly this is not the case [see section 2.2]. However, alternative conceptual approaches suffer from the same limiting factors. As new vehicle tests are performed, and as re-analysis of past vehicle tests continues, progress towards a representative fleet will be accomplished. In fact, the technology group approach provides greatest potential for correct representation when representative samples are not provided.

- *Fleet Characteristics Module Conceptual Completeness*

The 'conceptual completeness' of the vehicle characteristics approach is fairly good, all operating automobiles are considered. However, data limitations severely hamper comprehensive development. By using a region's entire registration dataset, a comprehensive view of the region's vehicle characteristics is possible. Geocoding errors and decoding errors result in a significant loss in data [Bachman, 1998]. Problems with completeness are resolved by developing distributions of technology groups instead of frequencies. While lost data have bias and cannot be fully represented, the use of distributions provides a 'best guess' given the data limitations.

- *Fleet Characteristics Enterprise Awareness*

The vehicle characteristics can be tied to other users of spatially-resolved fleet descriptions because the locational parameters are defined before the fleet is segmented into emission-specific technology groups. This allows the individual vehicle characteristics to be available for other analyses. Inclusion of technology groups from two separate modeling approaches directly results in added flexibility and openness for the users.

### **3.4.3. Vehicle Activity**

As mentioned previously, the core prognostic capability of the model rests on the ability of travel demand forecasting models to accurately predict regional travel. The emission-important vehicle activity estimates provided by the regional travel models are: the number and location of peak hour (or daily) trip origins, road segment volumes, and road segment average speeds. Important activity not provided by current models are; temporal travel behavior and modal (idle, cruise, acceleration, and deceleration) operations. As indicated in the background chapter, the average speed estimates can be very poor. Therefore, it is the role of the vehicle activity modules to transfer usable travel model information into the modeling environment, and develop estimates of the missing important parameters.

#### **3.4.3.1. *Engine Start Activity***

Engine starts are equivalent to trip origins determined by the trip generation component of travel demand forecasting models. Travel demand forecasting models divide an urban region into traffic analysis zones (TAZs). The TAZs represent a spatial unit for aggregating socioeconomic data and resulting trip generation estimates. The designation of a TAZ should be based on homogeneous socioeconomic characteristics, reducing the variability of the trip estimates. However, many urban areas use zonal definitions based on cadastral (property) boundaries or US Census boundaries. TAZs are usually large (2-5 sq. km) due to the original objectives of the travel demand models (major infrastructure investments). Unless the TAZs can be disaggregated to smaller zones, the TAZ structure will determine the spatial resolution.

Estimates of trip generation are made for each TAZ for a variety of trip purposes. Trip purposes usually include trip production and attraction estimates of home-based work (to and from the workplace), home-based shopping, home-based school, home-based other, and non-home based. While these trips are estimated to begin or end in certain TAZs, the trip type definitions imply that they can be tied to land use. For example, a home-based-work trip consists of a trip originating from home (residential) going to work (non-residential), or a trip originating from work going home. Likewise, home-based shopping trips imply trips to or from a commercial land use.

The US Census Bureau maintains zonal databases developed for the decennial census. The smallest zonal designation is a block, usually an area bounded by roads or other lineal features (cadastral, hydrologic, etc.). At the census block level, 1990 estimates of the number of households are available. While the estimates are out-of-date, they can possibly provide clues to housing density within the TAZ and land use designations. This information can be used to further spatially disaggregate trips originating from residential areas.

By having good land use data and socioeconomic data, various trips can be disaggregated to smaller zones. Even if the land use designations are as broad as “residential” and “non-residential”, the spatial resolution of trip generation estimates can be improved, allowing an improved spatial resolution for engine start estimates.

#### **3.4.3.2. *Intra-zonal Running Exhaust Activity***

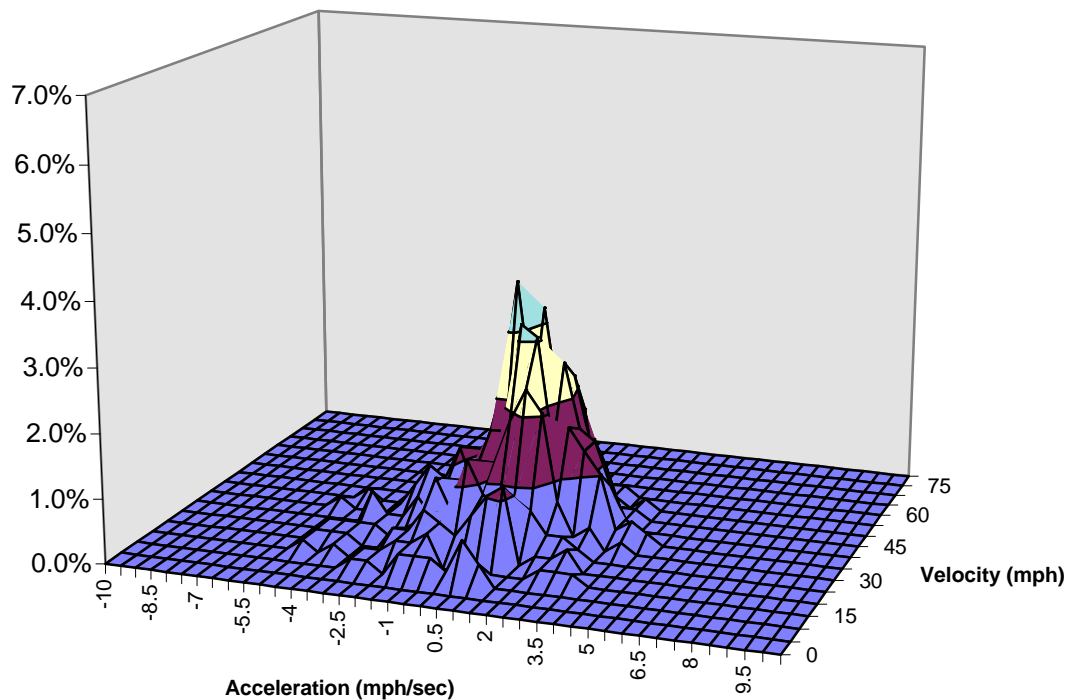
The road network used by travel demand forecasting models usually consists of major roads only. Travel on other roads is either not considered or predicted on an aggregate zonal (TAZ) basis. A key variable in predicting running exhaust emissions is the amount of travel time (preferably broken down by operating mode) because the longer a vehicle is operating, the more pollutant is produced. Travel times for intra-zonal trips (and inter-zonal travel off the major roads) are unaccounted for, other than looking at the size of the zone. However, information exists that allows the development of travel time estimates using the previously mentioned disaggregate trip generation estimates, a digital road network, and spatial analysis tools provided by the geographic information systems (GIS).

Many GISs provide tools that allow the determination of the shortest network path between two points. The disaggregated trip generation estimates provide a trip origin location. The closest intersection of local roads and major roads provides a destination location, representing the point during the vehicle trip when the travel demand models have assigned trips to the network. The shortest network path between the two points provides an estimate of the travel distance. Averaging all the distances within a TAZ provides an estimate of the typical intra-zonal and inter-zonal travel distance that occurs before vehicles reach the modeled network. Assuming an average speed for the local road travel provides an estimate of the average travel time. Although the strategy described above is crude and unvalidated, the method is better than the alternatives of leaving the estimates out, or assuming travel times based on zone area.

#### **3.4.3.3. *Modal Activity***

Modal activity is a vehicle activity characterized by cruise, idle, acceleration or deceleration operation. Research has clearly identified that modal operation is a better indicator of emission rates than average speed [see section 2.2]. Determining regional modal operation is not possible using current travel demand forecasting models alone. All travel models can provide is road volume ( $\pm 15\%$ ) and average speed ( $\pm 30\%$ ). Because the accuracy of the average speed is poor, it should not be used in emission rate evaluation. However, the average speed could be accurate enough to determine differences in levels of service (LOS) E and F where volume to capacity (v/c) ratios approach or surpass 1.

Research by Grant et al. is attempting to characterize speed and acceleration profiles (Watson plots) by collecting data on major roads around Atlanta with a Laser Rangefinder [Grant, 1996]. The research has produced results for freeway and ramp sections by grade, LOS, and vehicle type. Using these results, speed and acceleration profiles can be identified for prevailing conditions predicted by the travel demand forecasting model. An example of a speed / acceleration profile is provided in Figure 3.3. The figure shows a graph where the x variable is speed in 5 mph increments (0-80), the y variable is acceleration in 0.5 mph/sec (+10 to -10) increments, and the z variable is the fraction of activity.



**Figure 3.3 - Speed / Acceleration Profile, Interstate Ramp, LOS D**

#### **3.4.3.4. Road Grade**

The impacts of road grade on emissions are included in the model design. Road grade affects vehicle emissions by impacting the load on the engine. Gravity exerts a force on a vehicle that must be counteracted to maintain a constant speed. Road grade is not included in mandated emission models because tests on the actual effects have not been completed and because metropolitan areas do not maintain spatially defined road grade estimates. Although grade impacts on emission rates are being researched,

results are not available at this time. However, the effects of acceleration on emissions have been quantified. Therefore, the secondary effects of grade on acceleration can be included in the conceptual design.

The effects of grade on acceleration can be quantified by the equation:

$$\text{Acceleration}_{(\text{mph/sec})} = 22.15_{(\text{mph/sec})} \times (\text{Gradient}_{(\text{road})})$$

where 22.15<sub>(mph/sec)</sub> represents acceleration due to gravity. For example, a vehicle wishing to maintain a constant speed along a 5% road grade must accelerate 1.11<sub>(mph/sec)</sub> to counteract deceleration due to gravity.

Road grade data, while not currently comprehensively available for urban areas, is information that can be collected using global positioning systems (GPS) [Awuah-Baffour, 1997]. Given the expected importance of grade in affecting running exhaust emission rates, it is likely that the new GPS strategies will be employed by metropolitan areas in the next few years.

Including vehicle activity impacts resulting from road grade, even if not fully developed, provides an important step in emission model development. Strategies that are used for developing connections between road grade data and other road characteristics will act as guides in the development of future load-based models.

#### **3.4.3.5. Temporal Variability**

The temporal variability in estimates of vehicle activity is highly inaccurate because it relies on traditional travel demand forecasting models [see section 2.3.1]. Travel demand forecasting models are designed and operated to predict peak hour travel or daily travel. These are the primary temporal aggregation levels used by transportation planners and traffic engineers. The Travel Model Improvement Program (TMIP) administered by the US Department of Transportation is researching strategies for travel models to better predict activity during off-peak hours.

The ability of the research model to predict hourly emissions will rely heavily on accurate vehicle activity measures throughout the day. Until progress is made in the TMIP research, this emissions model will be unable to accurately incorporate off-peak travel. However, an intermediate step between existing and future models is possible. Many MPOs have developed estimates of hourly or subhourly travel demand factors based on travel survey data. These regional factors by trip type can be used to disaggregate daily trip generation into hourly intervals. Data on the variability of road volume throughout the day are available from departments of transportation for many major roads. Although average speed cannot be predicted to determine LOS F during

off-peak hours, volume-to-capacity ratios provide sufficient information for selection of appropriate speed and acceleration profiles.

Although these steps only provide an alternative to temporally comprehensive travel modeling, they allow the research model framework to prepare for future improved estimates of travel activity.

- *Vehicle Activity Conceptual Correctness*

The ‘conceptual correctness’ of the vehicle activity refers to the accurate portrayal of the actions of an aggregate group of vehicles, or, in other words, the ability to predict the distribution of activity in a zone or on a link. The largest source of error comes from the travel demand forecasting model. Heavy reliance on the model transfers errors in trip generation estimates, road volumes, and road speeds to the emission models. The research approach attempts to lessen the impact of these errors by using modal activity measures when possible, and by disaggregating trip origins to appropriate land uses. The use of temporal factors causes substantial error in the activity estimates because the factors are used region-wide and lack spatial variability. While better than current practice, the approach results in significant problems with accurate representation of emission-specific vehicle activity.

- *Vehicle Activity Conceptual Completeness*

The ‘conceptual completeness’ of the representation of emission-specific vehicle activity refers to the ability of the modeling approach to capture all of the important activities. The largest gap in the completeness of the representation occurs on non-highway or ramp roads. Few speed and acceleration profiles are available for major and minor arterials. An enhancement to the travel model is a linkage between trip origins and the major road network. Minor road shortest paths allow vehicle activity to be estimated between the zonal-based starts and the lineal-based running exhaust. Further, inadequate representation of activity around signalized and unsignalized intersections may cause the exclusion of a large source of emission-specific vehicle activity. Until other data are available that can help determine these operations, the completeness of the vehicle activity estimates will be poor.

- *Vehicle Activity Enterprise Awareness*

The estimates of vehicle activity can be tied to other enterprises through the zonal aggregations and road network. The road network maintains a variety of locational parameters including street address and travel model identifiers.

#### **3.4.4. Facility Emissions**

Facilities are divided into zones and lines corresponding to the previously mentioned emission modes of engine starts and running exhaust (respectively). Facility

estimates are used to allocate emission production to those vector spatial data structures currently used by transportation planners. By tying emission production estimates to facilities, tasks regarding research, reporting, validation, or control strategy development are made easier.

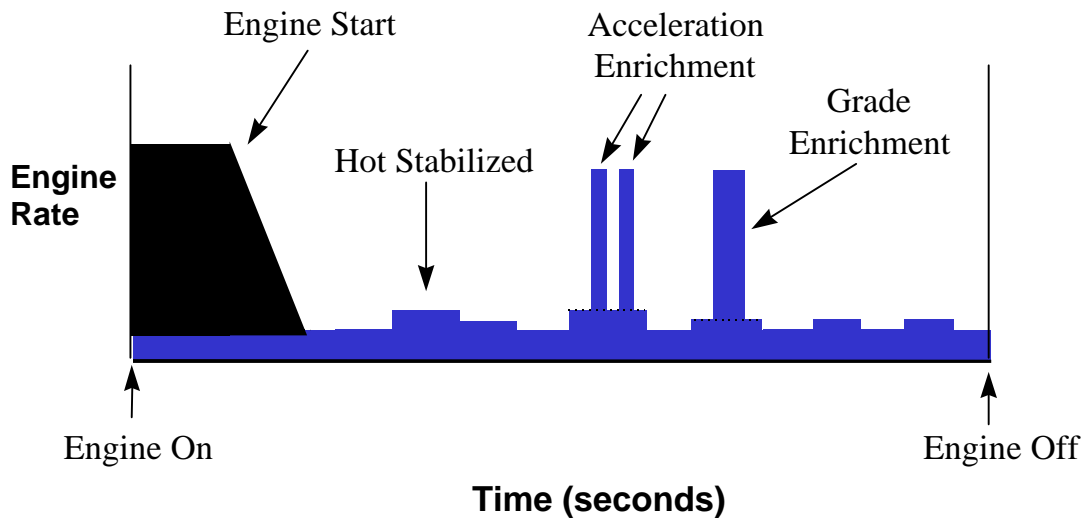
#### **3.4.4.1. *Engine Start Zonal Facility Estimates***

Zonal facilities include the zonal representations of TAZs, land use, and Census blocks. The model design allows for other zonal designations to be included, but only the three mentioned have been required. The zones have been included in the definition of facilities because they are used by planners to aggregate socioeconomic information. While running exhaust emissions occur within zones, they are better tied to modal activity that occurs on the road. Engine starts, however, occur at trip origins, generally characterized with point or zonal information.

Figure 3.4 schematically represents the portion of a vehicle's emission profile represented by zonal facilities. The exhaust engine start estimates are modeled as a 'puff' ( all start emissions allocated to the trip origin). While start emissions are actually dispersed through the network as a vehicle travels, research has not identified a strategy for correct spatial allocation. However, the role of this model is the study of emission production by automobiles, not air quality. It may be more useful for planners and/or researchers to have start emissions tied to the point of origin, thus allowing linkages to other zonal information.

Engine start emission rates are included in the research model based on results from the statistical model [see section 2.2.3]. Emissions in grams per start are estimated using the regression tree mentioned in sections 3.4.2.3 and 3.4.2.4. Six technology group trees exist for engine start emissions, all based on vehicle technology characteristics. Each emission estimate has established confidence bounds that can be translated back through the model to assess accuracy. The technology characteristics used in the tree process are listed below:

- *MY = Model Year*
- *EMM = Emission Control Equipment, 1-none, 2-oxi, 3-cat, 4-oxi&cat*
- *FINJ = Fuel injection equipment, 1-port, 2-carb, 3-throt*
- *CID = Engine Size, Cubic Inch Displacement*
- *TWT = CERT test weight, lbs.*



**Figure 3.4 - Engine Start Emission Portion**

The resulting technology groups are mutually exclusive and listed below:

- *CO Normal:*
  1. MY < 1981, TWT < 3250
  2. MY < 1980, TWT ≥ 3250, TWT < 4375
  3. MY < 1980, TWT ≥ 4375, CID < 351
  4. MY < 1980, TWT < 4375, CID ≥ 351
  5. MY ≥ 1980, TWT ≥ 3250
  6. MY = 1981, TWT ≥ 3688, CID < 131
  7. MY = 1981, TWT < 2938, CID < 131
  8. MY = 1981, TWT ≥ 2938, TWT < 3688, CID < 131
  9. MY ≥ 1982, MY < 1987, TWT < 3688
  10. MY ≥ 1982, MY < 1987, TWT ≥ 3688
  11. MY ≥ 1987

- *CO High:*

1. CID < 116, FINJ < 2
2. CID < 116, FINJ >= 2
3. CID >= 116, CID < 134
4. CID >= 134, CID < 258, FINJ < 2, MY < 1986
5. CID >= 134, CID < 258, FINJ = 2, TWT < 3563, MY < 1986
6. CID >= 134, CID < 258, FINJ = 2, TWT >= 3563, MY < 1986
7. CID >= 134, CID < 258, FINJ = 2, MY >= 1986
8. CID >= 134, CID < 258, FINJ >= 3
9. CID >= 134, CID >= 258

- *HC Normal:*

1. MY < 1980, TWT < 4125, CID < 154
2. MY < 1980, TWT < 4125, CID >= 154, CID < 241
3. MY < 1980, TWT >= 4125, CID >= 241
4. MY < 1978, TWT >= 4125
5. MY >= 1978, MY < 1980, TWT >= 4125
6. MY >= 1980, EMM < 4, CID < 171
7. MY >= 1980, EMM < 4, CID >= 171
8. MY >= 1980, MY < 1988, EMM >= 4, CID < 98
9. MY >= 1980, MY < 1988, EMM >= 4, CID >= 98, CID < 102
10. MY >= 1980, MY < 1988, EMM >= 4, CID >= 102
11. MY >= 1988

- *HC High:*

1. MY < 1980
2. MY >= 1980, FINJ < 3, CID < 196
3. MY >= 1980, FINJ < 3, CID >= 196, CID < 258, MY < 1987
4. MY >= 1980, FINJ < 3, CID >= 196, CID < 258, MY >= 1987
5. MY >= 1980, FINJ < 3, CID >= 258, MY < 1983
6. MY >= 1980, FINJ < 3, CID >= 258, MY >= 1983
7. MY >= 1980, FINJ >= 3, TWT < 2688
8. MY >= 1980, FINJ >= 3, TWT >= 2688, MY < 1988, CID < 192, TWT < 3063
9. MY >= 1980, FINJ >= 3, TWT >= 3063, MY < 1988, CID < 192
10. MY >= 1980, FINJ >= 3, TWT >= 2688, MY < 1988, CID >= 192
11. MY >= 1980, FINJ >= 3

- *NOx Normal:*

1. EMM < 3
2. EMM >= 3, EMM < 4, CID < 230
3. EMM >= 3, EMM < 4, CID >= 230, CID < 245, FINJ < 2
4. EMM >= 3, EMM < 4, CID >= 230, CID < 245, FINJ >= 2

5. EMM  $\geq$  3, EMM  $<$  4, CID  $\geq$  245
6. EMM  $\geq$  4, CID  $<$  122
7. EMM  $\geq$  4, CID  $\geq$  122, CID  $<$  138
8. EMM  $\geq$  4, CID  $\geq$  138, CID  $<$  146
9. EMM  $\geq$  4, CID  $\geq$  146, CID  $<$  152
10. EMM  $\geq$  4, CID  $\geq$  152, CID  $<$  213
11. EMM  $\geq$  4, CID  $\geq$  213, CID  $<$  288
12. EMM  $\geq$  4, CID  $\geq$  288

- *NOx High:*

1. EMM  $<$  3, CID  $<$  334, MY  $<$  1980
2. EMM  $<$  3, CID  $<$  334, MY  $\geq$  1980
3. EMM  $<$  3, CID  $>$  334
4. EMM  $\geq$  3, CID  $<$  137, MY  $<$  1987
5. EMM  $\geq$  3, CID  $<$  137, MY  $\geq$  1987
6. EMM  $\geq$  3, CID  $\geq$  137, CID  $<$  152
7. EMM  $\geq$  3, CID  $\geq$  152, CID  $<$  230
8. EMM  $\geq$  3, CID  $\geq$  230, CID  $<$  232
9. EMM  $\geq$  3, CID  $\geq$  232

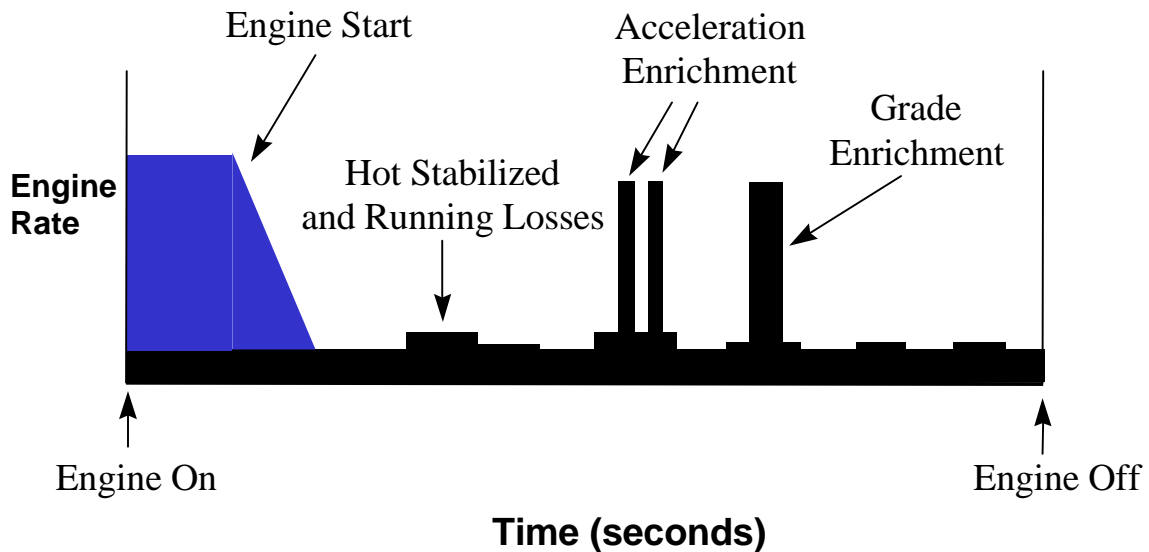
Zonal estimates of fleet characteristics are divided into the previous technology groups. Each technology group fraction is multiplied by the number of trip origins that occur in the zone. The resulting number of trip origins by technology group is multiplied by the associated gram per start emission rate. The resulting emissions of CO, HC, and NOx are reported for a typical weekday (Tuesday - Thursday) on an hourly basis. The typical weekday limitation is a result of the travel demand modeling process, as few models predict weekend or Friday travel.

#### **3.4.4.2. Minor Road Zonal Facility Estimates**

Minor road zones [see section 3.4.1.2] are used to spatially represent the portion of running exhaust emissions that occur between the trip origin and the roads modeled by the travel demand forecasting network. Available minor road vehicle activity information is limited because it is not explicitly modeled in the travel forecasting process, and there are no existing measures of modal activity available for local roads. Lower traffic densities and lower average speeds suggest that the actual portion of running exhaust emissions occurring on local roads may be small. However, little evidence is available to draw conclusions about the impacts of local road driving. This limitation forces a scaled back version of local road emissions modeling.

#### 3.4.4.3. Lineal Facility Estimates

Lineal facilities are roads that are modeled in the travel demand forecasting model. On-road fleet distributions and predicted traffic flow parameters are used to generate road segment specific estimates of CO, HC, and NOx. Figure 3.5 shows the portion of the emission spectrum represented by linear features. Minor road running exhaust emissions and major road running exhaust emissions estimate the same pollutant and, combined, predict total running exhaust emissions. Network characteristics determine the amount of the running exhaust portion that is allocated to minor zones.



**Figure 3.5 - Running Exhaust Emission Portion**

Emission rates for running exhaust come from two sources, the statistical approach used for start emissions, and the SCF approach, used by currently mandated models. The purpose for including both approaches is to allow user flexibility and to provide a platform for comparison. Vehicle activity measures for both emission rate approaches come from the same source, although different variables are needed. The SCF approach needs average speed, while the statistical approach uses a variety of

other modal parameters. The same vehicles are aggregated and used for the estimated fleet distribution, although different technology group definitions exist.

The technology groups (see section 3.4.4.2 for variable definitions) for the statistical model were determined similarly to those described in the engine start section. The definitions are as follows:

- *CO Normal:*

1.  $EMM < 4, MY < 1979$
2.  $EMM < 4, MY \geq 1979$
3.  $EMM \geq 4, CID < 146, MY < 1979$
4.  $EMM \geq 4, CID < 146, MY \geq 1979$
5.  $EMM \geq 4, CID \geq 146, MY < 1979$
6.  $EMM \geq 4, CID \geq 146, MY \geq 1979, MY < 1985$
7.  $EMM \geq 4, CID \geq 146, MY \geq 1985, MY < 1987$
8.  $EMM \geq 4, CID \geq 146, MY \geq 1987$

- *CO High*

1.  $EMM < 4, FINJ < 3, TWT < 3375$
2.  $EMM < 4, FINJ < 3, TWT \geq 3375$
3.  $EMM < 4, FINJ \geq 3$
4.  $EMM \geq 4, TWT < 3313$
5.  $EMM \geq 4, TWT \geq 3313$

- *HC Normal:*

1.  $MY < 1985$
2.  $MY \geq 1985, MY < 1987, TWT < 3188$
3.  $MY \geq 1985, MY < 1987, TWT \geq 3188$
4.  $MY \geq 1987, MY < 1990, CID < 143$
5.  $MY \geq 1987, MY < 1990, CID \geq 143, CID < 196$
6.  $MY \geq 1987, MY < 1990, CID \geq 196$
7.  $MY \geq 1990, TWT < 3563, CID < 143$
8.  $MY \geq 1990, TWT < 3563, CID \geq 143, CID < 186$
9.  $MY \geq 1990, TWT \geq 3563, CID < 143$
10.  $MY \geq 1990, TWT \geq 3563, CID \geq 143, CID < 186$
11.  $MY \geq 1990, CID \geq 186, CID < 196$
12.  $MY \geq 1990, CID \geq 196$

- *HC High:*

1.  $MY < 1984, EMM < 4$
2.  $MY \geq 1984, MY < 1985, EMM < 4$
3.  $MY \geq 1985, MY < 1988, EMM < 4$

4. MY < 1985, EMM >= 4, TWT < 4000
5. MY < 1985, EMM >= 4, TWT >= 4000
6. MY >= 1985, MY < 1988, EMM >= 4
7. MY >= 1988

- *NOx Normal:*

1. MY < 1989, CID < 154
2. MY >= 1989, MY < 1994, CID < 154
3. MY >= 1994, MY < 1995, CID < 154
4. MY >= 1995, CID < 154
5. MY < 1994, CID >= 154
6. MY >= 1994, MY < 1995, CID >= 154
7. MY >= 1995, CID >= 154

Each engine start technology group will have a gram per start emission rate. Running exhaust emissions rates depend on modal operation. The regression tree used to determine running exhaust emissions rates includes modal parameters. The modal parameters are:

- *AVGSPD* - The average speeding miles per hour.
- *PKE>X* - The fraction of activity with positive kinetic energy ( $\text{speed} \times \text{acceleration}$ ) greater than  $X \text{ mph}^2/\text{sec}$ .
- *POW>X* - The fraction of activity with power ( $\text{speed}^2 \times \text{acceleration}$ ) greater than  $X \text{ mph}^3/\text{sec}$ .
- *ACC>X* - The fraction of activity with acceleration greater than  $X \text{ mph}/\text{sec}$ .
- *DEC>X* - The fraction of activity with deceleration greater than  $X \text{ mph}/\text{sec}$ .
- *CRZ>X* - The fraction of activity with zero acceleration and speed greater than  $x \text{ mph}$ .
- *IDLE* - The fraction of the activity with zero acceleration and zero speed.

For each road segment and each hour, modal variables are determined. Road segment-specific technology groups and modal variables are combined to develop the fraction of activity with specific emission rates (grams per second). Total hourly travel time is calculated and segmented by the fraction of the vehicles with each emission rate.

The SCF emission rate approach uses a 'look-up' table created from running MOBILE5a for a series of model year distributions and average speed distributions. The vehicle characteristics developed for the model include model year as a variable, allowing the creation of model year technology groups. The vehicle activity component of the model allows the estimate of average speed to be available for every

road segment and every hour. By running MOBILE5a for each combination of average speed and model year distribution, gram per second emission rates will be developed.

- *Facility Emissions Estimates Conceptual Correctness*

Conceptual correctness for facility emission estimates is the ability of the approach to accurately estimate actual emission production. Sources of inaccuracies come from three places; the quality of the input data, the model's manipulation of the data, and the errors in the development of emission rates. The quality of the input data and the ability of the model to manipulate them into a usable form result in large errors. All of the problems mentioned in the previous sections culminate in substantial error. Errors in the development of the emission rates are the result of incomprehensive test datasets and unrepresentative operating profiles. Emission rates have confidence bounds associated with them, allowing some measure of the variability of the errors. As in the representation of vehicle activity, the conceptual correctness of the facility emission rates is defined by the aggregate level of the estimate. While the accurate assessment of a single vehicle may be poor, the aggregate estimate may prove better. The SCF emission rates have been strongly criticized as being insensitive to important activity. Thus, the SCF emission rates have less 'correctness' than the statistically based approach that includes modal parameters.

- *Facility Emissions Estimates Conceptual Completeness*

The 'conceptual completeness' of the exhaust emissions estimates is fairly strong. All automobiles will fit into engine start, running exhaust, and SCF technology groups. All operating vehicles will fall in the speed and acceleration profile identified for the particular conditions. The only source of incomplete representation of emission rates is due to problems that have already been mentioned regarding the emission test dataset and the cycle test design. All represented vehicles can be assigned an emission rate.

- *Facility Emissions Estimates Syntactic Correctness and Completeness*

The communication of the facility emission estimation component relies on clear definitions and visual representation. The terminology used to describe the technology groups and other input parameters can be clearly communicated if the user is given concise explanations of all the parameters. The spatial component of the facility estimates can be clearly communicated using GIS and by using the input data spatial structures.

- *Facility Emissions Estimates Enterprise Awareness*

The estimates have strong 'enterprise awareness' due to the ability of the estimates to be aggregated to any of the original input spatial structures (TAZs, Census, etc.). If needed, the estimates can also be presented as emissions per unit of distance or area to aid in translating the other locational parameters. This step is

improved through the translation to raster structure in the next section; however, the base information needed is available at this level.

### **3.4.5. Emissions Inventory**

The role of the emissions inventory module is to prepare the facility-based emission estimates for input into gridded photochemical models. An important component of the entire modeling process is the ability to aggregate estimates to a user-defined grid cell size. The most efficient technique for accomplishing this task is to convert the engine start, minor road running exhaust, major road running exhaust, and major road SCF emission estimates to raster data structures. Once in a raster structure, developing gridded estimates for inventory reporting is fairly easy. Conversion of the data from vector to raster is a tool available in many of the larger GIS software tools. After conversion, total mobile source emission estimates are calculated for the entire area. Engine starts, minor road running exhaust, and major road running exhaust emissions are used to develop totals.

The tools available in the GIS for conversion make some assumptions about the vector data that may not be desired. Problems occur with direct conversion, especially for linear structure. Straight conversion is possible, but grid cells take on the value of the largest feature, or largest portion within its boundaries. For a linear feature, this means that all cells that represent the line will have the same emission value or rate. However, the line can bisect the cell at any point, resulting in variations in the cell's ability to properly identify the portion of the road that falls within its boundaries. Similar issues occur with polygons along their boundaries. The smaller the grid cell, the lesser the problem. However, gridded inventories require grid sizes as large as 4-5 square kilometers, much larger than the anticipated zonal and lineal structures.

One way around the problem is to intersect the zonal and lineal structures with a vector grid. Once intersected, all emission values falling within the grid cell boundaries can be weighted by area or length and summed. The resulting vector grid data are then converted to raster cells with cell sizes equivalent to the vector grid size.

The final raster datasets are individual 'layers' of each pollutant by hour and emission 'mode' (totals, engine starts, etc.). Tools available in the GIS allow for the development of special visualization interfaces that can create and query two and three dimensional images of the various databases.

- *Gridded Emissions Conceptual Correctness*

For the final module, 'conceptual correctness' refers to the ability to maintain data integrity while converting from raster to vector. The deterministic, vector approach to developing gridded estimates ensures that few errors are introduced during

the process. There are also problems with grid cells that fall on the boundaries of the study area. Cells that overlap study area boundaries will only have values for those emissions in the study area, causing the cell value to be underestimated. As long as cells lie completely within the boundaries, this is not a problem. The accuracy of the gridded estimates is a function of all of the previously discussed problems.

- *Gridded Emissions Conceptual Completeness*

The completeness of the process is only limited when grid structures do not encompass or lie within the entire study area.

- *Gridded Emissions Enterprise Awareness*

Aggregating and rasterizing the emission estimates allow the storage and communication of specific emission production intensities. Previously, a zone would represent only emission produced by ‘zonal’ information like trip generation estimates. With the gridded emissions, every location has an estimate of the total emissions as well as the specific modes. The flexible locational parameters allow the estimates to be translated to other areas.

### **3.5. Conclusion**

This chapter defined the parameters and quality assurance measures for the design of a comprehensive modal emissions model. Initially, a detailed list of issues to be addressed by the model was compiled from background research. This list of issues led to a model design that incorporated those parameters while maintaining a flexible, comprehensive and accurate account of the science being modeled. In developing the design, GIS became a powerful tool for data preparation, storage, and analysis.